**1. Introduction**

This project analyses retail transaction data to uncover customer behaviour, product performance, and revenue trends. Using **SQL** as the primary tool, the analysis focuses on delivering actionable insights that a business can directly use for decision-making. The objective is not only to explore the dataset but also to showcase advanced SQL techniques, ranging from aggregation queries to window functions and RFM-based customer segmentation.

**2. Methodology**

* **Database Creation**: Structured tables were created for customers, orders, order\_items, products, and payments. Data was imported using MySQL Import Wizard.
* **SQL Queries**: Multiple queries were executed to analyze customer demographics, product demand, revenue contributions, payment patterns, and customer value tiers.
* **Advanced SQL**: Window functions, subqueries, and case-based logic were applied for ranking, running totals, contribution percentages, and segmentation.

I have drawn Insights by asking business questions.

We are popular among which cities , what are the cities that contribute to most of our customers?

-- Total customers and top 10 cities by customer count

SELECT count(customer\_id) as Customer\_count ,city

FROM customers

GROUP BY city

ORDER BY Customer\_count DESC

LIMIT 10 ;

A screenshot of a computer

AI-generated content may be incorrect.

Our business is doing great in the East James and is really performing consistently in majority of cities.

How much products we sell category wise ?

-- Unique product categories + count per category

SELECT DISTINCT category , count(product\_id) as count\_product

FROM products

GROUP BY category

ORDER BY count\_product;

A screenshot of a computer

AI-generated content may be incorrect.

We sell most products from the Beauty category followed by fitness and home décor , but we must also know if we just put them on sell or the category is actually popular among our customers and has got good sell .

So we would answer this question to get better understanding ,

Which is most ordered category of the product ?

-- Most ordered category of product

SELECT

p.category,

count(o.quantity) as quantity\_ordered

FROM

products as p

JOIN order\_items as o

ON p.product\_id = o.product\_id

GROUP BY

p.category

ORDER BY

quantity\_ordered desc;

A screenshot of a table

AI-generated content may be incorrect.

So the picture is clear, We sell the quantity of products of the categories in accordance of their demand and popularity among the customers

How much do our business earn monthly , what is the trend and is their consistency ?

-- Total revenue per month

SELECT

monthname(order\_date) as Month,

sum(total\_amount) as monthy\_revenue

FROM orders

GROUP BY Month

ORDER BY monthy\_revenue ;

A screenshot of a graph

AI-generated content may be incorrect.

There is no visible outliers or any spike or downs in our monthly sales , we are at extend performing very consistent . There is slight decrease in the revenue in July we must look deeply into underlying reasons and so goes the same for highest revenue during march , what went so well.

We saw which cities has the large customer base so we would like to see,

which 10 cities reap the highest number of orders ?

-- Top 10 Cities by Total Orders

SELECT

c.city,

count(o.order\_id) as total\_orders

FROM

customers as c

JOIN orders as o ON

c.customer\_id = o.customer\_id

GROUP BY c.city

ORDER BY total\_orders desc

LIMIT 10

A screenshot of a table

AI-generated content may be incorrect.

We got the idea of the top 10 cities wherein the customer most order from so we much manage the delivery network to keep the customers happy.

What is the recent trend in market , which is the most frequently order product ?

-- Most Frequently Purchased Products

SELECT p.product\_name ,

p.category,

count(o.quantity) as quantity\_ordered

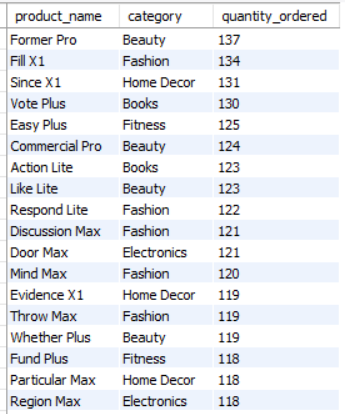
FROM products as p

join order\_items as o ON

p.product\_id = o.product\_id

GROUP BY p.product\_name , p.category

ORDER BY quantity\_ordered desc



The list signifies that we are in right direction of the selling the right products as earlier we had drawn conclusion of popularity of beauty and fashion.

Which category contributes most to our overall revenue ?

-- Revenue by Category

SELECT p.category,

sum(o.total\_amount) as revenue

FROM products as p

JOIN order\_items as ot

ON p.product\_id = ot.product\_id

JOIN orders as o

ON o.order\_id = ot.order\_id

GROUP BY p.category

ORDER BY revenue desc;

A screenshot of a computer

AI-generated content may be incorrect.

The beauty and fitness is the biggest contributor of the revenue. But the fashion products as also among most ordered yet the revenue is not to the mark , concluding that cart value for the quantity ordered is low.

What is the contribution of the all payment methods in the total of successful payments done?

-- Payment Method Split (Success Only)

SELECT payment\_method ,

count(payment\_id) as count\_payments

FROM

payments

WHERE payment\_status = 'success'

GROUP BY payment\_method;

A screenshot of a computer

AI-generated content may be incorrect.

Introduction of the wallet payment option is a success.

How is our recent performance , what is the order count last 30 days ?

-- Daily Order Count Trend (Last 30 Days)

SELECT order\_date as day ,

count(order\_id) as total\_orders

FROM orders

WHERE order\_date >= curdate() - interval 30 day

GROUP BY day

ORDER BY day

A screenshot of a computer

AI-generated content may be incorrect.

17-33 is out range of the quantity ordered each day in the past 30 days . The bracket is not too varied , we have considerable consistent performance

Who are our star and most value customers ?

-- Customers who spent more than ₹50,000 total : High Value Customers

SELECT c.customer\_id ,

c.name ,

sum(o.total\_amount) as spendings

FROM customers as c

JOIN orders as o

ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_id,

c.name

HAVING spendings >50000

ORDER BY spendings desc;



We got a huge list of high value customers , we can contact them personally and introduce a loyalty program for them.

Some Advance queries

Who are our high monetary value customers by rank?

-- Rank customers by their total spending from highest to lowest.

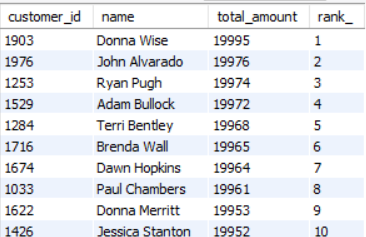
SELECT c.customer\_id , c.name , o.total\_amount,

dense\_rank()OVER(ORDER BY o.total\_amount desc ) as rank\_

FROM customers as c

JOIN orders as o

ON c.customer\_id = o.customer\_id ;



The glimpse of top 10 customers who contribute to our monetary value , hence introducing loyalty program for them will be great.

How much had we earned till now cumulated?

-- Running Monthly Revenue Total

WITH MonthlyRevenue AS

( SELECT DATE\_FORMAT(order\_date, '%Y-%m') as revenue\_month ,

SUM(total\_amount) as monthly\_revenue

FROM orders

GROUP BY revenue\_month)

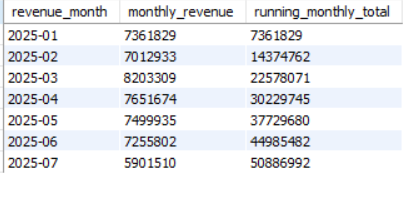
SELECT revenue\_month , monthly\_revenue ,

sum(monthly\_revenue)OVER( ORDER BY revenue\_month

ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW)

AS running\_monthly\_total

FROM MonthlyRevenue;



This Signifies that we are making some good money.

How much does each product contribute to our total revenue?

-- Product Sales % Contribution

WITH productsales AS

(SELECT p.product\_name AS product , SUM(oi.quantity \* oi.unit\_price) as total\_sales , sum(oi.quantity) as quantity\_ordered

FROM products as p JOIN

order\_items as oi ON oi.product\_id = p.product\_id

JOIN orders as o ON o.order\_id = oi.order\_id

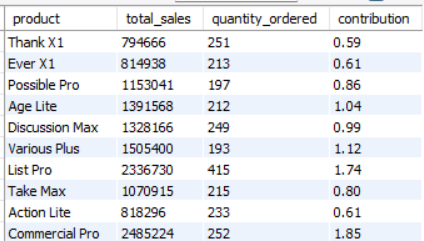
GROUP BY product )

SELECT product , total\_sales , quantity\_ordered ,

ROUND(total\_sales \* 100.0 / SUM(total\_sales) OVER (),2)

AS contribution

FROM productsales ;



The percentage distribution of each product to out total.

Who are the customers who bought above average quantity from us?

-- Customers Who Bought Above-Average Quantity

WITH cust\_qty AS (

SELECT o.customer\_id, SUM(oi.quantity) AS total\_qty

FROM orders as o

JOIN order\_items as oi ON o.order\_id = oi.order\_id

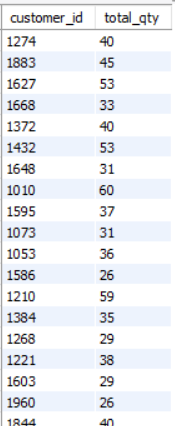
GROUP BY o.customer\_id

)

SELECT customer\_id, total\_qty

FROM cust\_qty

WHERE total\_qty > (SELECT AVG(total\_qty) FROM cust\_qty);



WHAT ARE OUR TOP 3 SELLING CATEGORIES AND THEIR CONTRIBUTION TO TOTAL ?

-- Top 3 Categories by Revenue

WITH contribution AS

(SELECT p.category , sum(o.total\_amount) AS revenue , sum(quantity) AS n\_quantity

FROM products as p JOIN order\_items as oi ON oi.product\_id = p.product\_id

JOIN orders as o ON o.order\_id = oi.order\_id

GROUP BY p.category

ORDER BY revenue desc )

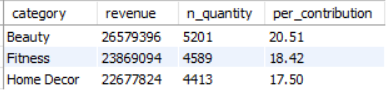
SELECT category , revenue , n\_quantity ,

ROUND(revenue \*100 / sum(revenue) over(),2 ) as per\_contribution

FROM contribution

ORDER BY revenue desc

LIMIT 3



Almost 60% of our revenue is generated by the top 3 selling categories of our business.

How is segmentation of our customers by their spending tier ?

-- Classify Customers by Spend Tier

SELECT c.name , c.customer\_id , sum(o.total\_amount) as spendings ,

CASE

WHEN sum(o.total\_amount) > 50000 THEN 'high value'

WHEN sum(o.total\_amount)BETWEEN 20000 AND 50000 THEN 'medium value'

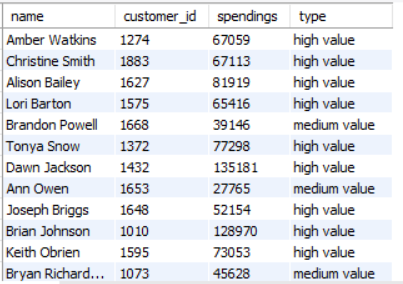
WHEN sum(o.total\_amount) < 20000 THEN 'low value'

END AS type

FROM customers as c JOIN orders as o

ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_id , c.name



Now according to this segmentation we can frame the policies , loyalty program and discounts and offer trends customized for each type .

RFM ANALYSIS

WITH rfm\_base AS (

SELECT

c.customer\_id,

c.name,

MAX(o.order\_date) AS last\_purchase,

COUNT(DISTINCT o.order\_id) AS frequency,

SUM(o.total\_amount) AS monetary

FROM customers c

JOIN orders o ON c.customer\_id = o.customer\_id

GROUP BY c.customer\_id, c.name

),

rfm\_scores AS (

SELECT

customer\_id,

name,

DATEDIFF(CURDATE(), last\_purchase) AS recency,

frequency,

monetary,

-- Recency score

CASE

WHEN DATEDIFF(CURDATE(), last\_purchase) <= 30 THEN 5

WHEN DATEDIFF(CURDATE(), last\_purchase) <= 90 THEN 4

WHEN DATEDIFF(CURDATE(), last\_purchase) <= 180 THEN 3

WHEN DATEDIFF(CURDATE(), last\_purchase) <= 365 THEN 2

ELSE 1

END AS recency\_score,

-- Frequency score

CASE

WHEN frequency >= 200 THEN 5

WHEN frequency >= 100 THEN 4

WHEN frequency >= 50 THEN 3

WHEN frequency >= 10 THEN 2

ELSE 1

END AS frequency\_score,

-- Monetary score

CASE

WHEN monetary >= 10000 THEN 5

WHEN monetary >= 5000 THEN 4

WHEN monetary >= 1000 THEN 3

WHEN monetary >= 500 THEN 2

ELSE 1

END AS monetary\_score

FROM rfm\_base

)

SELECT \*,

CASE

WHEN recency\_score >= 4 AND frequency\_score >= 4 AND monetary\_score >= 4

THEN 'Champions'

WHEN frequency\_score >= 4 AND recency\_score >= 3

THEN 'Loyal Customers'

WHEN monetary\_score >= 4 AND recency\_score >= 3

THEN 'Big Spenders'

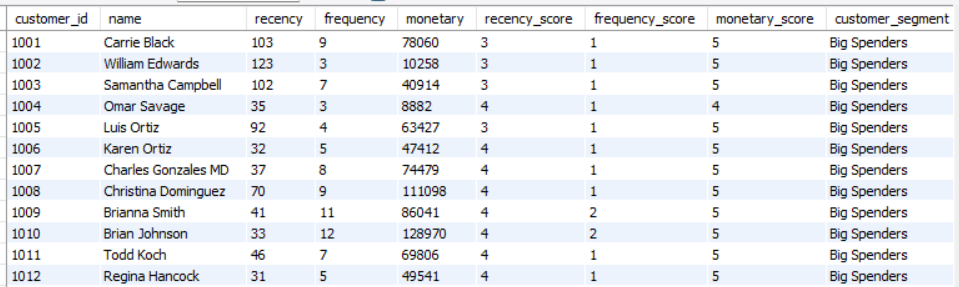
WHEN recency\_score <= 2 AND frequency\_score >= 3

THEN 'At Risk'

ELSE 'Others'

END AS customer\_segment

FROM rfm\_scores;



**4 . Key Findings and Insights**

**1. Customer Distribution & Behavior**

* **A significant share of customers and orders are concentrated in the top 10 cities, especially East James, showing geographic concentration of demand.**
* **Customers naturally segment into High Value (>₹50K), Medium Value (₹20–50K), and Low Value (<₹20K) spend tiers.**
* **RFM analysis highlights that a small group of “Champions” and “Loyal Customers” drive recurring revenue, while a chunk of “At Risk” customers need reactivation campaigns.**

**Implication: Growth depends on (a) strengthening delivery & marketing in top-performing cities and (b) designing tiered loyalty programs to retain and upgrade customers.**

**2. Product & Category Performance**

* **Beauty and Fitness categories dominate both demand and revenue contribution.**
* **Fashion has high order volume but low revenue, indicating low average order value (AOV) — a margin concern.**
* **Around 60% of revenue comes from top 3 categories, and a few “star products” generate disproportionate sales.**
* **Some categories are underperforming with low demand, making them candidates for discounting or phasing out.**

**Implication: Prioritize star categories & products for inventory and promotions, while rationalizing weaker categories. Push premium SKUs in Fashion to improve profitability.**

**3. Revenue & Sales Trends**

* **Monthly revenues are steady overall, with a clear spike in March (likely seasonal/campaign-driven) and a dip in July.**
* **Daily sales (last 30 days) are relatively stable (17–33 orders per day) with no major volatility.**

**Implication: Revenue base is consistent, but growth levers lie in (a) replicating successful campaigns like March, and (b) running targeted weekday/seasonal promotions to smooth out dips.**

**4. Payment Behaviour**

* **Digital payments (wallets, UPI, cards) dominate successful transactions, while COD lags and has higher risk of failure/returns.**
* **Introduction of wallet payments has been well-received.**

**Implication: Encourage digital payments further with cashbacks/discounts and work to reduce COD failure/leakage through better delivery & fraud checks.**

**5. Customer Value Concentration**

* **Top 10–20% of customers contribute a majority of revenue (Pareto effect).**
* **High-value customers are prime candidates for VIP loyalty benefits and personalized engagement.**

**Implication: Focus on retention over acquisition for revenue stability — a small base of high-value customers sustains business growth.**

**6. Strategic Takeaways**

* **Customer Management: Loyalty rewards for Champions, upselling for Medium tier, reactivation offers for At Risk.**
* **Product Strategy: Protect and grow Beauty & Fitness, improve Fashion margins, phase out non-performers.**
* **Revenue Strategy: Benchmark growth vs. targets; replicate successful sales periods; use daily sales tracking to boost weaker days.**
* **Payment Optimization: Promote digital-first strategy while reducing COD risks.**
* **Geographic Strategy: Invest in logistics, faster delivery, and city-specific campaigns in top-performing regions.**

**5. Conclusion**

This project demonstrates the power of SQL in extracting actionable insights from retail data. By combining fundamental queries with advanced techniques like window functions and RFM segmentation, the analysis provides a **360-degree view** of customers, products, and revenue streams. These insights not only strengthen day-to-day operational decisions but also inform strategic planning for growth, customer retention, and product development.